

02-18-00

A

02/17/00



Jc760 U.S. PTO

Please type a plus sign (+) inside this box → ☒
 PTO/SB/05 (4/88)  
 Approved for use through 09/30/2000. OMB 0651-0032  
 Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

<b>UTILITY PATENT APPLICATION TRANSMITTAL</b> <i>(Only for new nonprovisional applications under 37 C.F.R. § 1.53(b))</i>	Attorney Docket No.	Y0999-195 (8728-281)
	First Inventor or Application Identifier	Ittycheriah
	Title	SYSTEMS AND METHODS FOR PROCESSING...
	Express Mail Label No.	EL433927819US

<b>APPLICATION ELEMENTS</b> <i>See MPEP chapter 600 concerning utility patent application contents.</i>	<b>ADDRESS TO:</b> Assistant Commissioner for Patents Box Patent Application Washington, DC 20231
--	---

1. <input checked="" type="checkbox"/> Fee Transmittal Form (e.g., PTO/SB/17) <i>(Submit an original and a duplicate for fee processing)</i> 2. <input checked="" type="checkbox"/> Specification [Total Pages 40] <i>(preferred arrangement set forth below)</i> - Descriptive title of the invention - Cross References to Related Applications - Statement Regarding Fed sponsored R & D - Reference to Microfiche Appendix - Background of the invention - Brief Summary of the invention - Brief Description of the Drawings (if filed) - Detailed Description - Claim(s) - Abstract of the Disclosure 3. <input checked="" type="checkbox"/> Drawing(s) (35 U.S.C. 113) [Total Sheets 6] 4. Oath or Declaration [Total Pages ] a. <input type="checkbox"/> Newly executed (original or copy) b. <input type="checkbox"/> Copy from a prior application (37 C.F.R. § 1.63(d)) <i>(for continuation/divisional with Box 16 completed)</i> i. <input type="checkbox"/> DELETION OF INVENTOR(S) Signed statement attached deleting inventor(s) named in the prior application, see 37 C.F.R. §§ 1.63(d)(2) and 1.33(b).	5. <input type="checkbox"/> Microfiche Computer Program (Appendix) 6. Nucleotide and/or Amino Acid Sequence Submission <i>(if applicable, all necessary)</i> a. <input type="checkbox"/> Computer Readable Copy b. <input type="checkbox"/> Paper Copy (identical to computer copy) c. <input type="checkbox"/> Statement verifying identity of above copies
---	---

<b>ACCOMPANYING APPLICATION PARTS</b> 7. <input type="checkbox"/> Assignment Papers (cover sheet & document(s)) 8. <input type="checkbox"/> 37 C.F.R. § 3.73(b) Statement <input type="checkbox"/> Power of Attorney <i>(when there is an assignee)</i> 9. <input type="checkbox"/> English Translation Document (if applicable) 10. <input type="checkbox"/> Information Disclosure Statement (IDS)/PTO-1449 <input type="checkbox"/> Copies of IDS Citations 11. <input type="checkbox"/> Preliminary Amendment 12. <input checked="" type="checkbox"/> Return Receipt Postcard (MPEP 503) <i>(Should be specifically itemized)</i> * Small Entity <input type="checkbox"/> Statement filed in prior application, Status still proper and desired 13. <input type="checkbox"/> Statement(s) <i>(PTO/SB/09-12)</i> 14. <input type="checkbox"/> Certified Copy of Priority Document(s) <i>(if foreign priority is claimed)</i> 15. <input checked="" type="checkbox"/> Other: ASSOCIATE POWER OF ATTORNEY	
---	--

16. If a CONTINUING APPLICATION, check appropriate box, and supply the requisite information below and in a preliminary amendment:

☐ Continuation ☐ Divisional ☐ Continuation-in-part (CIP) of prior application No: \_\_\_\_\_

Prior application information: Examiner \_\_\_\_\_ Group / Art Unit: \_\_\_\_\_

For CONTINUATION or DIVISIONAL APPS only: The entire disclosure of the prior application, from which an oath or declaration is supplied under Box 4b, is considered a part of the disclosure of the accompanying continuation or divisional application and is hereby incorporated by reference. The incorporation can only be relied upon when a portion has been inadvertently omitted from the submitted application parts.

17. CORRESPONDENCE ADDRESS

☐ Customer Number or Bar Code Label \_\_\_\_\_ or ☐ Correspondence address below

(Insert Customer No. or Attach bar code label here)

Name	Frank V. DeRosa				
Address	F. Chau & Associates, LLP 1900 Hempstead Turnpike, Suite 501				
City	East Meadow	State	New York	Zip Code	11554
Country	USA	Telephone	516-357-0091	Fax	516-357-0092

Name (Print/Type)	Frank V. DeRosa	Registration No. (Attorney/Agent)	43,584
Signature	<i>Frank V. DeRosa</i>	Date	2/17/00

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Box Patent Application, Washington, DC 20231

Assistant Commissioner for Patents  
Washington, D.C. 20231  
Sir:

ATTORNEY DOCKET: Y0999-195 (8728-281)  
Date: February 17, 2000  
Express Mail Label: EL433927819US  
Date of Deposit: February 17, 2000

Transmitted herewith for filing is the Patent Application of:

Inventors: Abraham P. Ittycheriah, Stephane H. Maes, Jan Sedivy

For: SYSTEMS AND METHODS FOR PROCESSING AUDIO USING MULTIPLE SPEECH TECHNOLOGIES

[X] Priority of U.S. Provisional Application No. 60/136,671 filed on May 28, 1999 is claimed under 35 U.S.C. 119(e)

Enclosed are: [X] 32 sheets of specification; [X] 1 sheet(s) of Abstract; [X] 7 sheet(s) of claims; [X] 6 sheet(s) of drawing(s);

[ ] An assignment of the invention to International Business Machines Corporation with Recordation Form.

[ ] Declaration and Power of Attorney.

[ ] A certified copy of a \_\_\_\_\_ application, from which priority under Title 35 USC §119 is claimed.

[X] Associate Power of Attorney.

The filing fee has been calculated as shown below:

(Col. 1)		(Col. 2)	OTHER THAN A SMALL ENTITY	
FOR:	NO. FILED	NO. EXTRA	RATE	FEE
BASIC FEE				\$690.00
TOTAL CLAIMS	27 -20 =	7	X \$18 =	126.00
INDEP CLAIMS	3 -3 =	0	X \$78 =	0
MULTIPLE DEPENDENT CLAIMS PRESENTED			+ 260 =	
			TOTAL	\$ 816.00

If the difference in Col. 1 is less than zero, enter "0" in Col. 2.

[ ] A check in the amount of \$\_\_\_\_\_ to cover the [ ] filing fee(s), [ ] recording fee is enclosed.

[X] Please charge my Deposit Account No. 50-0510/IBM (Yorktown Heights) in the amount of \$816.00.

[X] The Commissioner is hereby authorized to charge payment of the following fees associated with this communication or credit any overpayment to Deposit Account No. 50-0510/IBM (Yorktown Heights). A duplicate copy of this sheet is enclosed.

[X] Any additional filing fees required under 37 CFR 1.16.

[X] Any patent application processing fees under 35 CFR 1.17.

Respectfully submitted,

By:

Frank V. DeRosa  
Frank V. DeRosa  
Registration No. 43,584

Please address all  
correspondence to:

F. CHAU & ASSOCIATES, LLP  
1900 Hempstead Tpke., Suite 501  
East Meadow, NY 11554  
Tel: (516) 357-0091  
Fax: (516) 357-0092

Attorney for:  
IBM Corporation  
Intellectual Property Law Dept.  
P.O. Box 218  
Yorktown Heights, NY 10598

**CERTIFICATION UNDER 37 C.F.R. §1.10**

I hereby certify that this Application transmittal and documents referred to as enclosed are being deposited with the United States Postal Service on this date February 17, 2000 in an envelope as "Express Mail Post Office to Addressee" Mailing Label Number EL433927819US addressed to: Assistant Commissioner for Patents, Box Patent Application, Washington, D.C. 20231.

Frank V. DeRosa  
Frank V. DeRosa

Attorney Docket No.: YO999-195 (8728-281)

U.S. Patent Application:

Title:               SYSTEMS AND METHODS FOR PROCESSING AUDIO  
                      USING MULTIPLE SPEECH TECHNOLOGIES

Inventors:        Abraham P. Ittycheriah, Stephane H. Maes,  
                      Jan Sedivy

Assignee:         International Business Machines Corporation

Filed:             February 17, 2000

F. Chau & Associates, LLP  
1900 Hempstead Turnpike, Suite 501  
East Meadow, NY 11554  
Tel: (516) 357-0091  
Fax: (516) 357-0092

**SYSTEMS AND METHODS FOR PROCESSING AUDIO  
USING MULTIPLE SPEECH TECHNOLOGIES**

**Cross-Reference to Related Application**

This application is based on provisional application  
5 U.S. Serial No. 60/136,671 filed on May 28, 1999.

**BACKGROUND**

**1. Technical Field:**

10 The present invention relates generally to data processing systems and, more particularly, to systems and methods for providing managed sharing of audio data between multiple speech technologies.

**2. Description of Related Art:**

15 Currently, there are many speech/audio processing systems in which audio data or processed speech data is stored in buffers for consumption and further processing by speech engines. The conventional systems, however, typically do not include mechanism for properly balancing the load on engines and managing the consumption of data from the buffers. For instance, in the area of telephony  
20 DSP (digital signal processing) cards, conventional systems

include a hardware based TDM (time-division multiplexed) bus which carries speech data to single or multiple destinations. This architecture requires the use of dedicated chips to transport the signal as well as physical cards. These systems do not provide intelligent routing of the speech stream which may cause the speech stream to be transmitted twice to the same host.

In addition, in the area of embedded architectures, the currently existing systems have very limited capabilities. For example, these embedded systems typically operate by having an audio subsystem assigned temporarily to a specific conversational engine until the audio subsystem is released either by the engine, the controlling application or the underlying operating system.

Furthermore, conventional sound card systems, in general, capture an audio waveform and store the waveform in digitized form in a buffer. Typically, these systems are configured such that only one application will be consuming the content of the buffer at a given time. In specific cases, however, where an utterance is shared between different engines one of the following methods may be used. One method includes a hardware implementation of multiple parallel buffers on the sound card to which multiple engines could connect. Although such soundcard configuration is not

commercially available at the present time, a hardware implementation would require adding the necessary circuitry to route the data stream to the aforementioned buffers. Such a system would not provide intelligent management of the consumption or tailoring of the systems resources according to the evolution of the speech sharing. With another method, a single buffer through one engine may be used which thereafter saves the utterance in a logged file for consumption by the other engines. These engines receive the file name and path information as handle to the data. Again, intelligent management of the data consumption in such an architecture is nonexistent.

Furthermore, with systems that generate output speech (playback or output from TTS), the output is typically sent to an output buffer that is consumed by a D/A converter of the audio subsystem. Such an approach typically does not provide management the output consumption, especially in conjunction with the input resource requirements when operating in a full duplex mode.

Accordingly, a system and method that provides intelligent routing and sharing of speech data for consumption by multiple engines operating in a given speech system is highly desirable.

## SUMMARY OF THE INVENTION

The present invention is directed to system and methods for sharing speech data associated with the same utterance between multiple speech technologies. In one aspect of the present invention, a system for sharing data between multiple consumers (or data splitting system) comprises a first queue for storing data; a plurality of consumers each sharing the data stored in the first queue; and a scheduler for managing the storage of the data in the first queue and the consumption of the data in the first queue by each of the plurality of consumers.

In another aspect of the present invention, the system comprises a plurality of queues and plurality of consumers. The consumers may include speech engines such as feature extraction engines, speech decoding engines, and speaker identification/verification engines, as well as data compression and decompression engines. The consumers will register their data requirements and priority requests with the scheduler. The scheduler assigns each of the plurality of consumers to one or more of the plurality of queues based on the registered data requirements. In this manner, the sharing of audio data (i.e., audio splitting) can occur at different stages in an I/O processing chain by, e.g., distributing digitized waveforms between different consuming

engines and distributing features obtained at several stages of processing of the audio stream.

In yet another aspect, for each queue in the system, the scheduler maintains an IN pointer associated with the data source that feeds the queue and one OUT pointer for each of the plurality of consumers assigned to the queue, so as to manage the flow of the data in and out of the queue. Using these pointers, the scheduler can determine how much of the shared data has been read/not read by each of the consumers. The scheduler will prioritize data consumption of the queue based on an amount of unread data of each of the of consumers assigned to the queue.

The present invention may be implemented on various platforms. For instance, in one embodiment, an audio splitting system according to the present invention may be implemented in an embedded engine. In another embodiment, an audio splitting system may be implemented in a telephony system. In yet another embodiment, an audio splitting system may be implemented in an audio playback/processing system.

These and other aspects, features and advantages of the present invention will be described and become apparent from the following detailed description of preferred embodiments,



which is to be read in connection with the accompanying drawings.

#### **BRIEF DESCRIPTION OF THE DRAWINGS**

Fig. 1 is a block diagram of a system for processing audio data using multiple speech technologies according to one embodiment of the present invention;

Fig. 2 is a diagram of a system/method for processing audio data using multiple speech technologies according to another embodiment of the present invention;

Fig. 3 is a diagram of an embedded speech recognition engine in accordance with a preferred embodiment of the present invention, that may be employed in the system of Fig. 2;

Fig. 4 is a diagram which illustrates a scheduling method in accordance with one aspect of the present invention, which may be implemented in the system of Fig. 3;

Fig. 5 is a diagram of a system/method for processing audio data using multiple speech technologies according to another embodiment of the present invention; and

Fig. 6 is a diagram illustrating a method for managing a queue according to one aspect of the present invention.

## DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

It is to be understood that the exemplary system and methods described herein may be implemented in various forms of hardware, software, firmware, special purpose processors, or a combination thereof. Preferably, the present invention is implemented in software as an application program tangibly embodied on a program storage device. The application program may be executed by any machine, device or platform comprising suitable architecture. It is to be further understood that, because some of the constituent system components and method steps depicted in the accompanying Figures are preferably implemented in software, the actual connections between the system components (or the process steps) may differ depending upon the manner in which the present invention is programmed. Given the teachings herein, one of ordinary skill in the related art will be able to contemplate these and similar implementations or configurations of the present invention.

Referring now to Fig. 1, a block diagram illustrates a system for supporting multiple speech processing technologies according to an embodiment of the present invention. In particular, the system of Fig. 1 illustrates a telephony system 10 according to an exemplary embodiment of the invention (which is preferably implemented on a

computer platform) for providing one or more telephony services. The telephony system 10 comprises a speech splitting subsystem 10a and a speech processing subsystem 10b having a plurality of engines for processing the audio data so as to provide a variety of speech services (e.g., speech recognition, speaker identification, etc.) In general, the speech splitting subsystem 10a is configured for duplicating an audio data stream (e.g., speech data) and allocating each of the duplicated audio streams to a corresponding engine in the speech processing subsystem 10b based on the particular speech services that are requested from, e.g., an application that is currently managing the telephone line. For example, if the application is a banking application, a request may be for a speech recognizer to recognize an account number and another request may be made for a speaker verification engine to verify the speaker that is using the account number. More specifically, the systems and methods disclosed in U.S. Patent No. 5,897,616, issued on April 27, 1999 to Kanevsky et al., entitled "Apparatus and Methods For Speaker Verification/Identification/Classification Employing Non-Acoustic and/or Acoustic Models and Databases" and U.S. Serial No. 08/788,471, filed on January 28, 1997, entitled "Text Independent Speaker Recognition for Transparent

Command Ambiguity Resolution And Continuous Access Control,"  
are typical applications where the present invention may be  
employed.

5 The telephony system 10 comprises a telephone card (not  
shown) to monitor a telephone line and detect and answer and  
incoming call. When an incoming call is answered, the  
telephony system 10 will record the incoming audio data  
(e.g., speech data) via audio module 11 and store the  
recorded audio in audio buffer 12. The audio module 11 is  
10 the portion of the application that is responsible for  
capturing and converting (and possible processing) the  
speech and transmitting the speech data to single or  
multiple destinations.

15 The audio splitting subsystem 10a comprises one or more  
telephony speech managers (TSM) 13 (i.e., control points)  
which share a single recorded audio stream. Each TSM object  
13 is preferably defined as a class to provide a specific  
API (application program interface) for accessing certain  
telephony services. In particular, each TSM object 13  
20 serves as an API to a corresponding TSM processing subsystem  
(TSMP) 17a and 17b of the speech processing system 10a.  
Each TSM object 13 is the control side (client) for making  
requests to a specific technology or speech engine based on  
the requests for particular speech services from the

managing application. In particular, each TSM object 13 will make requests to the Audio object 11 in response to particular requests from the managing application for transmitting speech data to a particular speech processing engine.

Each TSMP 17a and 17b of the speech processing subsystem 10b comprises a TSM process 18a and 18b, respectively. Each TSM process 18a and 18b acts as a "server" to process requests from corresponding TSM objects 13 on the client side TSM (e.g., to enable a vocabulary). In addition, each TSMP 17a and 17b comprises a SMAPI (speech manager application program interface) 19a and 19b (or any other conventional API that is suitable for the given application) which provides low-level interfaces to corresponding engines 21a and 21b in engine subsystems 20a and 20b.

The audio splitting subsystem 10a further comprises a mailbox module 14 connected to each TSM object 13. Each mailbox module 14 serves as a communication mechanism for the client and operates to find a particular engine requested by the client. More specifically, each mailbox module 14 may query a TSM router 16 to determine if there is a TSMP 17a and 17b associated with an engine 21a and 21b that is responsible for performing the requested speech

processing function (e.g., speech recognition, speaker identification etc.). It is to be understood that other suitable communication mechanisms and protocols (other than mail boxes) may be employed herein. The TSM Router process

5 16 maintains a list of servers that it owns in a table. The table also indicates the state of the engines, e.g., whether particular engines are allocated or free. When a request is received from a client side TSM object 13, the TSM Router 16 will scan its table to located the desired engine type. If

10 a requested engine is found, the TSM Router 16 will deliver a handle to that engine (i.e., the TSM router 16 returns the handle of the TSMP (e.g., 17a and 17b) that is responsible for performing the requested function). The appropriate connection will then automatically be established to the

15 server via, e.g., RPC (remote procedure call), sockets, windows messages, etc. The engine is then marked "busy" and is then freed by the TSM client side object 13 after the speech function is complete.

A socket audio module 15 (or "audio splitter" module)

20 is configured for duplicating the speech and transmitting the speech in real time to each engine subsystem 20a and 20b which are allocated for performing the requested speech services. Particularly, the socket audio module 15 utilizes any suitable standardized protocol such as TCP/IP

(transmission control protocol/internet protocol), RTP (realtime protocol), Voice over IP, or other streaming protocols, for transmitting the duplicated speech data to the desired engine subsystems 20a and 20b. Each engine subsystem 20a and 20b comprises an audio listener module 22a and 22b, respectively, which listens on, e.g., an TCP/IP socket to receive transmitted speech data having a corresponding protocol address.

It is to be appreciated that because a common protocol is utilized (i.e., all the recipients agree to accept a common format), the format of the transmitted speech data can be unconstrained. In addition, to reduce the network bandwidth and reduce the processing at each recipient, the duplicated data streams can be preprocessed (e.g., pre-emphasis, filtering, acoustic feature extraction, labeling, etc.).

It is to be understood that the speech splitting subsystem 10a may be network-connected to the speech processing subsystem 10b for providing remote speech processing services, wherein each engine is located on a single computer or server or multiple computers or servers. In addition, the entire speech processing subsystem 10b may reside on a single computer or server along with the speech splitting subsystem 10a. It is to be appreciated that if the

plurality of engines 21a and 21b reside on a common host,  
then only a single connection need be made to the remote  
host so as to ship the audio data once using only the  
minimal network bandwidth required. In this manner, a proxy  
5 process can be included on the second host whereby, instead  
of directly talking to the engines, the telephony system can  
talk to the proxy which would then duplicate the stream on  
the host.

It is to be appreciated that the audio splitting  
10 process described above with respect the telephony system  
can be implemented with other platforms, for example, at the  
level of an embedded engine. Referring to Fig. 2, a block  
diagram illustrates an exemplary system architecture  
according to another embodiment of the present invention for  
15 splitting audio and processing the audio using multiple  
speech processing technologies using an embedded speech  
engine (ESE). The system comprises an embedded speech  
engine (ESE) 200 having a kernel 201 (core engine) that  
implements and executes one or more algorithms 202 (e.g.,  
20 speech recognition, speaker identification algorithms) under  
the control of an API 203. The API 203 comprises a set of  
functions which enable the basic control of the kernel 201.  
The ESE 200 utilizes speech data files 204 (or  
conversational arguments) such as grammars, language models,



vocabulary files, voiceprints, etc., for decoding speech data from audio source 207.

The ESE 200 executes on top of a local operating system (or platform) having corresponding operating system functions 206. An application 205 (e.g., an audio processing program) which is managed by, and utilizes the resources of, the operating system, communicates with the kernel 201 via the API 203 when particular speech services are needed by the application 205. It is assumed that the ESE 200 is optimized for support by the CPU of the machine, device, and/or platform in which the ESE 200 executes and is delivered to the CPU in compiled object form. The information need for implementing runtime data structures (which are dependent on the given computer platform architecture) are obtained from an implementation module 208.

For the exemplary embedded application, the ESE 200 is preferably written as highly portable C/C++ code independently on the operating system. More specifically, the architecture for the exemplary embodiment comprises a plurality of abstracted functions for communicating with the environment in which the ESE 200 executes (e.g., access to an audio subsystem, memory allocations, etc.) Such OS functions 206 are typically provided by the underlying OS or

by the hardware implementation. As such, when the C/C++ code comprising the ESE 200 is ported to a given platform, only the abstraction layer needs to be ported and correctly linked to the services that are provided by the underlying OS and hardware.

An audio source 207 (i.e., an object (abstraction) that has the prescribed behavior) is controlled by the developer of audio hardware (e.g., soundcard). The ESE 200 comprises abstracted functions for communicating with the audio source 207 and drive the hardware. The developer of the audio source 207 provides the appropriate links (i.e., the actual audio objects) that allows the abstracted functions of the ESE 200 to drive the hardware and have the hardware provide the appropriate behavior.

The application 205 includes listeners 209 which are functions that receive calls from ESE 200. The listener functions 209 are typically called when the ESE 200 informs the application 205 about certain events. For example, when the ESE 200 recognizes a word, a "result" listener is called. The listeners 209 are part of the application 205 and may be readily added to the application 206.

Fig. 3, illustrates a detailed architecture of the ESE 200 of Fig. 2 according to an exemplary embodiment of the present invention, which may be employed for providing audio

splitting. The exemplary ESE 200 architecture comprises a PCM queue 210 which receives audio data from the audio source 207. The audio data is preferably pulse code modulated (PCM) audio data that is output from an analog-to-digital (A/D) converter included in the hardware associated with the audio source 207 (although it is to be understood that the audio data may be coded using other conventional techniques). The buffering of the audio data in the PCM queue 210 serves various purposes. For instance, the buffering of the audio data prevents data loss. The ESE 200 is preferably run on a CPU having suitable speed so to achieve a fast response time and realize small latency (e.g., a recognized command should be executed fast without considerable delay). The CPU may, nevertheless (especially at beginning of an utterance), execute slower than at the end of an utterance when the variance of choices is lower. To deal with latency and prevent data loss, the PCM queue 210 buffers the PCM audio data. In addition, as explained in detail below, such buffering allows the audio data to be utilized by other consumers (e.g., speech engines).

As illustrated in Fig. 3, there are additional queues located at different levels in the audio path. For instance, the exemplary ESE 200 of Fig. 3 comprises three main executive units: digital signal processor (DSP) 211, a

labeler 213 and a decoder 215, all of which perform conventional functions. A CEP queue 212 is connected to the output of the DSP 211 and to the input of the labeler 213. The DSP 211 processes blocks of the buffered PCM audio data by performing, e.g., a cepstral analysis and generating a plurality of cepstral feature vectors representing the PCM audio data (such as MEL cepstra and/or LPC cepstra feature vectors). The feature vectors for each block of processed PCM audio data are stored in the CEP (cepstra) queue 212. The labeler 213 receives the feature vectors from CEP queue 212 and performs, e.g., a vector quantization process to determine the speech models (included in the speech data files 204) that are closest to the feature vectors (using any suitable conventional distance measure) and then ranks the feature vectors based on determined distances to the speech models. A rank queue 214 stores the rank information and corresponding feature vectors to await further processing by the speech decoder 215.

A scheduler 216 (or "supervisor" or "task dispatcher") performs functions such as assigning CPU time for the tasks of the different executive units during real-time operation based on the priorities designated to each of the executive units. To keep the buffers (e.g., queues) as small as possible, the data is preferably buffered in a place where

the data bandwidth is the narrowest which, as shown in Fig. 3, is between the DSP 211 and the labeler 213. To achieve this, the DSP 211 should be ready for execution as soon as a new block of PCM audio data is received. Consequently, the DSP 211 is afforded the highest priority. In addition, each of the executing units must run on a particular platform faster than the longest allowed time for the A/D converter servicing. The PCM queue 210 should be large enough to store the PCM audio data between two consecutive calls of the audio source 207. The PCM audio data is then transferred from the PCM queue 210 when the DSP 211 is called. On the other hand, the labeler 213 and the decoder 215 are preferably given the same priority. When two or more units have the same priority, the scheduler 216 decides which unit will be run first based on the size of the data in queue (e.g., the task having more data pending in the queue for processing will be executed first). After execution, the executive units will return to the scheduler 216.

During the time between the switching of two executive units, a special function `peek_app_idle()` 217 is called. This function is a listener 209 function and is implemented by the application designer inside the application 205 (as explained with reference to Fig. 2). This function affords

the application 205 the ability to, e.g., service the GUI every time an executive unit is switched. Generally, the peek\_app\_idle() 217 function should return to the scheduler 216 fast enough to provide its service between switching of two executive units. This function is especially useful for configurations running without support of a multitasking OS. With the support of multi-threaded or multitasking system, however, this function is not needed.

It is to be appreciated that by providing different queues at different levels in the audio path as illustrated in the exemplary ESE architecture of Fig. 3, audio splitting can be achieved at any level in the audio path by allowing multiple engines or executing units (referred to as consumers) to be connected to and obtain data from a given queue. That is, each of the internal queues can be employed to serve as a source of audio data for multiple consumers. This concept is illustrated by the block diagram of Fig. 4, wherein the CEP queue 212 is shown as servicing multiple speech engines (consumers) including not only the labeler 213 (as illustrated in Fig. 3), but also other executive units such as a speaker identifier 218, a speaker verifier 219 and a cepstrum listener 220. It is to be understood that the cepstrum listener 220 represents any other consumer of the cepstra in the CEP queue 212 such as another queue, a

dictation engine, a compression engine, or a networked data stream to send the cepstra, e.g., in compressed format to a networked engine using the methods described in U.S. Patent No. 6,009,387 to Ramaswamy, et al., issued December 28, 1999, entitled "System and Method Of Compression/Decompressing a Speech Signal By Using Split Vector Quantization and Vector Quantization," and International Appl. No. PCT/US99/22925, filed October 1, 1999, entitled "System and Method For Providing Network Coordinated Conversational Services", both of which are commonly assigned and incorporated herein by reference.

Each of the consumers will register their data requirements with the scheduler 216 and the scheduler 216 will assign a given consumer to one or more queues based on such requirements. For instance, as shown in Fig. 4, the labeler 213, speaker identifier 218 and speaker verifier 219, for example, are assigned to the CEP queue for processing the buffered cepstra. In one embodiment, the scheduler 216 will accord the same priority to each consumer assigned to a given queue and each consumer will inform the scheduler of the status of its consumption of data in the queue.

It is to be appreciated that the concept depicted in Fig. 4 may be further extended in the embedded architecture

of Fig. 3 by connecting multiple consumers to the PCM queue 209 and/or the rank queue 214 to provide audio spitting at those stages in the audio path. For instance, audio spitting may occur at the level of the PCM input (output of the A/D converter of the audio subsystem) to support multiple ESEs obtaining data from the same PCM buffer. An audio splitting process may occur the output of the DSP 211, e.g., splitting the acoustic features (such as MEL cepstra) by connecting, for example, multiple labelers of different speech engines (different languages). Moreover, an audio splitting process may occur at the output of the labeler 213, e.g., splitting the identity of prototypes (e.g., Gaussians) that best match the features that are obtained, for example, via vector quantization (or possibly, hierarchically). In addition, the output of the decoder 215 may be buffered for supporting audio splitting to additional consumers.

Referring now to Fig. 5, a block diagram illustrates a audio splitting system for supporting multiple speech technologies according to an embodiment of the present invention. In particular, Fig. 5 illustrates a general architecture for sharing audio data between multiple consumers, which may be employed in various platforms. The audio splitting system comprises an audio subsystem 300 and



a queue 301 that receives audio data from the audio subsystem 300. On a first level, the audio splitting system includes a first consumer 303 and a second consumer 304 which share the audio data stored in queue 301. On a second level, a queue 305, connected to the output of consumer 303, services multiple consumers 306 and 307. On additional levels, the consumer 306 may be attached to one or more queues which service other consumers.

It is be appreciated by those skilled in the art that the system of Fig. 5 may readily be modified to provide various levels and hierarchical configurations of consumers and queues depending on the intended application. For example, using the above embedded example, queue 301 may be a PCM queue, the first consumer 303 may be a DSP that processes the audio data in queue 301 to generate Mel cepstra feature vectors that are stored in queue 305 and the second consumer 304 may be another DSP that processes the audio data in queue 301 to generate LPC cepstra feature vectors (which may also be stored in a queue depending on the application). In addition, the queue 305 (having the MEL cepstra feature vectors) can service one or more labelers (e.g., consumers 306 and 307). In another example, different requirements in terms of the signal properties (sampling rate, bandwidth, file format) may

justify an audio split after the partial capture of the PCM data in queue 301. In particular, the PCM audio data may be captured at 22kHz and split at the output of queue 301 for servicing consumer 304 which is an acoustic front-end working at 22kHz sampling rate to extract cepstral feature, as well as servicing consumer 303 which is a down-sampling stage for other engines working at 11 or 8 kHz data. The output of this down-sampler (consumer 303) may be stored in queue 305 and split between different consumers 306 and 307.

The audio splitting system of Fig. 5, further comprises a scheduler 302 (or supervisor) which is responsible for load management/system management. The scheduler may be implemented using conventional system managers known in the art that are suitable for the given application such as the system management solutions offered the commercially available "Tivoli" (an application that provides load balance management for distributed systems). As generally described above, the scheduler 302 monitors the flow of data consumption and storage in the queues 301 and 305 and provides registration and prioritization services for each of the consumers (303, 304, 306, 307) in the system. In addition, for distributed systems, the scheduler 302 will monitor the network delay/traffic when scheduling and managing the different queues and consumer processes so as

to ensure, for example, efficient flow of network data to various consumers that are network-connected to a given queue.

More specifically, each consumer (e.g., engine) will register information such as the source that it consumes and the characteristics of the features that it consumes (e.g., what type of feature vector, etc.) Each source type is deemed a queue that is consumed by the registered consumers. The scheduler 302 monitors the consumption of such queues and changes the priorities of the tasks to ensure appropriate consumption of the queues. It is to be further appreciated that during the registration process, a given consumer may also register a priority request as a function of the state of the dialog or importance of the function. Such priorities are typically selected by the application developer or by the operating system.

The scheduler 302 manages the entire process chain by managing the CPU time and resources between different consumers of the different queues. For example, the scheduler 302 either favors or slows the engines to balance the different queues. The priority afforded to a given consumer (e.g., queue) which is a source for other consumers is influenced by both the state of the consumption of the

queue and by the state of consumption of the queues that it feeds.

Referring now to Fig. 6, a diagram illustrates a method for managing a queue according to one aspect of the present invention. For each queue in the system depicted in Fig. 5, for example, the scheduler 302 maintains a list of pointers pointing to data in and data out of the queue. In particular, as shown in Fig. 6a, for queue 400 (which comprises an allocated block of memory extending from a beginning memory location L1 to an ending memory location L5), the scheduler 302 maintains an IN pointer 401 that points to the next available memory location of the queue for storing speech data that is input from the source which feeds the queue 400. In addition, for each registered consumer of the queue 400, the scheduler 302 maintains an OUT pointer, e.g., OUT pointer 402 for consumer 1 and OUT pointer 403 for consumer 2. Each OUT pointer 402 and 403 points to the memory location having the next speech data to be consumed by the corresponding consumer.

As illustrated in Fig. 6a, the state of the queue is such that speech data input from the source fills memory location L1 through L3. It is to be understood that the "filled space" block of the queue represents the speech data that is shared (or split) between the multiple consumers

(e.g., consumer 1 and consumer 2). Furthermore, the OUT pointer 403 points to memory location L2 (indicating that consumer 2 has consumed the shared speech data in memory locations L1 through L2) and the OUT pointer 402 points to location L1 (indicating that consumer 1 has not consumed any of the shared speech data). When speech data is written to or read from the queue 400, the scheduler 302 will check the OUT pointers of each consumer and the IN pointer from the source to determine how much space is available for incoming data and how much data has not yet been processed. For instance, in Fig. 6b, the state of the queue 400 is similar to the state of the queue 400 of Fig. 6a except that, as illustrated in Fig. 6b, consumer 1 has consumed the shared speech data in memory locations L1 through L4. Since, at this state both registered consumers have consumed the speech data in memory locations L1 through L4, this space is "freed" to accept new input speech data from the source. The exemplary queue 400 of Fig. 6 operates as a circular buffer. Accordingly, the scheduler 302 can determine the amount of available memory in the queue for writing data based on the IN pointer 401 and the last OUT pointer (which is OUT pointer 402 in the illustrative embodiment).

The scheduler 302 controls the data consumption in a manner so as to prevent the queue 400 from overflowing. As

indicated above, assuming that registered consumer 1 and registered consumer 2 of queue 400 have the same priority, the scheduler 302 will grant priority to consumer 1 in Fig. 6 since it has more unread data in queue 400 than consumer 2. Indeed, by prioritizing consumer 1, more memory in the queue 400 can be made available (i.e., the "freed space" block increases). The queue 400 is emptied when all the consumers have read the data in the queue.

The scheduler 302 manages queue consumption by prioritizing and slowing down the different consumers to balance the queue consumption and avoid queue overflow. For instance, in a multiple port system, for example the scheduler 302 may reduce or increase the amount of engines instantiated per machine to balance the consumption of the queues (i.e. favoring the slowest consumers and slowing down the fast consumer and the producers earlier in the chain). This can also be done artificially (although much less efficient) by slowing down the network traffic towards the fast consumers. Furthermore, in a load balancing/system management topology, the slower consumers can simply receive more CPU, clock cycles similarly to the approach followed for the embedded engine embodiment discussed above. In addition, the scheduler can manage queue consumption in the distributed topology by monitoring the network traffic to

determine any possible delays in data flow between, e.g., a given consumer and its registered consumers that are remotely located on another machine or device.

When a given queue threatens to overflow, different mechanisms may be applied. For instance, multiple queues may be instantiated when a queue is almost full to store the additional input data and allow all consumers to obtain the necessary data before the data is released. Alternatively, based on the requirements of the consumers, the scheduler 302 may simply allow the data that was not collected by all the consumers to be lost after a given time period. In addition, when the priority of all the consumers in the entire system are properly set, the system designer can make the queue having the smallest bit rate the largest buffer. This buffer can then readily store its data when the system lags behind in real-time operation.

It is to be appreciated that by using the same software and/or hardware implementation throughout the system for implementing the buffers, the scheduler 302 and each registered consumer in the audio splitting system of Fig. 5 may be "re-programmed" by one of its consumers further down the chain. This allows a new conversational engine or function to be dynamically loaded in the existing audio management architecture. For example, assume that a given

consumer needs to consume another type of acoustic feature. A new DSP process may be programmed to consume the PCM input and generate the new acoustic feature type, which is then provided to the consumer. In another example, a DSP may be re-programmed to switch from producing one type of feature (e.g., for a speech recognition engine) to producing another type of acoustic feature (e.g., for another speech recognition task or for a particular speaker recognition engine, etc.) A "customization routine" to perform such programming can be loaded, for example, via an applet or an ActiveX component (or whatever other acceptable programming method is supported by the platform).

It is to be further appreciated that as indicated above, the system of Fig. 5 may be implemented as a networked architecture wherein certain consumers are network-connected to other groups of consumers on remote systems/platforms. In such case the scheduler 302 will distribute local priorities to each distributed system and monitor the network delay/traffic so as to maintain appropriate flow of data to each consumer. The transmission of audio data in compressed format between each of the distributed systems/consumers may be implemented using the compression and communication techniques disclosed in the



above-incorporated U.S. Patent No. 6,009,387 and  
International Appl. No. PCT/US99/22925.

It is to be understood that the present invention may  
be extended to platforms or systems other than the  
5 illustrative platforms/systems described herein. Such  
platforms/systems may include for instance, speech  
recognition systems, speaker recognition systems, utterance  
verification systems, speech biometric systems, natural  
language understanding systems, mood recognition systems,  
10 data segmentation systems (e.g., segmentation bases on  
speaker changes, environmental changes, etc.), data storage  
systems, data compression systems and communication systems  
for distribution of speech to networked distributed  
applications.

15 One platform in which the implementation of the audio  
splitting system of Fig. 5 is particularly useful is for  
providing management of a PC audio subsystem. Indeed, with  
conventional audio hardware implementation or key OS  
implementation, access to the audio stream is typically  
20 performed by using reserved memory blocks (e.g. 4 blocks of  
4K of memory) that are sequentially emptied by any engine  
that is operatively connected to the audio stream. In  
accordance with the present invention, a PCM queue may be  
employed at the output of the A/D converter of a PC audio

system (e.g., soundcard) rather than allocating fixed memory buffers (e.g., the 4 buffers) as is done conventionally. In this manner, the scheduler can manage the consumption of the output PCM queue as well as the writing data to the queue by changing the priorities of the source engines outputting data to the queue and the consuming engines. In addition, the only memory that needs to be allocated in accordance with the present invention is the memory that is needed to maintain the queue data at any given moment (as opposed to allocating the 4 buffers).

Another advantage of the present invention is that the hardware as well as the software, at each point of the chain (e.g., the DSP/labeler/decoder) or for any engine or technology utilizes the same architecture. As such, many processes at each stage can be operated in parallel.

Although illustrative embodiments of the present invention have been described herein with reference to the accompanying drawings, it is to be understood that the present invention is not limited to those precise embodiments, and that various other changes and modifications may be affected therein by one skilled in the art without departing from the scope or spirit of the invention. All such changes and modifications are intended

Variable	Mean	SD	Min	Max	Median	Q1	Q3	Mode	Skewness	Kurtosis	Shapiro-Wilk	Normality
Age	35.2	12.5	18	65	32	28	38	35	0.15	2.1	0.98	Normal
Gender	1.2	0.4	1	2	1	1	1	1	0.05	0.1	0.95	Normal
Marital Status	2.1	0.8	1	3	2	1	3	2	0.12	1.8	0.97	Normal
Education	15.8	2.5	10	20	16	15	17	16	0.08	0.5	0.99	Normal
Income	1200	300	500	2000	1100	900	1300	1000	0.18	2.5	0.96	Normal
Occupation	1.5	0.5	1	3	1	1	2	1	0.03	0.2	0.99	Normal
Health Status	2.5	0.5	1	3	2	2	2	2	0.01	0.1	0.99	Normal
Stress Level	3.2	1.0	1	5	3	2	4	3	0.10	1.5	0.98	Normal
Life Satisfaction	4.1	0.8	2	5	4	3	5	4	0.05	0.8	0.99	Normal
Work-Life Balance	3.8	0.9	2	5	4	3	5	4	0.02	0.3	0.99	Normal
Family Support	4.5	0.7	3	5	4	4	5	4	0.01	0.1	0.99	Normal
Community Involvement	2.8	0.6	1	4	3	2	4	3	0.04	0.6	0.99	Normal
Personal Growth	3.5	0.8	2	5	3	3	4	3	0.06	0.9	0.99	Normal
Overall Well-being	4.2	0.7	3	5	4	4	5	4	0.01	0.1	0.99	Normal

**WHAT IS CLAIMED IS:**

1. A system for sharing data between multiple consumers, comprising:

a first queue for storing data;

5 a plurality of consumers each sharing the data stored in the first queue; and

a scheduler for managing the storage of the data in the first queue and the consumption of the data in the first queue by each of the plurality of consumers.

10 2. The system of claim 1, further comprising a data source for outputting the data stored in the first queue, wherein the data output of the data source is controlled by the scheduler.

15 3. The system of claim 2, wherein the data source comprises a consumer that consumes data stored in a second queue.

4. The system of claim 2, wherein each of the plurality of consumers performs a registration process with the scheduler.

5. The system of claim 4, wherein the registration process between a consumer and the scheduler involves registering data requirements and priority requests of the consumer.

5 6. The system of claim 5, wherein the scheduler assigns each of the plurality of consumers to the first queue based on the registered data requirements.

10 7. The system of claim 2, wherein the scheduler maintains an IN pointer for the data source and one OUT pointer for each of the plurality of consumers to manage the flow of the data in and out of the first queue.

15 8. The system of claim 1, wherein the scheduler prioritizes data consumption of the first queue based on an amount of unread data of each of the plurality of consumers.

9. The system of claim 1, wherein the system is implemented in an embedded engine.

10. The system of claim 1, wherein the system is implemented in a telephony system.

11. The system of claim 1, wherein the first queue stores audio data output from an audio system.

12. The system of claim 1, wherein the first queue and the plurality of consumers are distributed over a network,  
5 and wherein the scheduler monitors data traffic over the network to manage and schedule data communication between the first queue and the plurality of consumers over the network.

13. The system of claim 1, wherein the system is  
10 dynamically re-programmed by one of the plurality of consumers or a new consumer.

14. The system of claim 1, wherein the system is dynamically re-programmed to include one of a conversational engine or a conversational function.

15 15. The system of claim 1, wherein the system is dynamically re-programmed by a procedural object.

16. A method for sharing data between multiple consumers, comprising the steps of:

storing data in a first queue;

sharing the data in the first queue between a plurality of consumers; and

managing the storage of the data in the first queue and the consumption of the data in the first queue by each of the plurality of consumers.

17. The method of claim 16, wherein the step of sharing comprises the steps of:

registering data requirements of each of the plurality of consumers; and

assigning the plurality of consumers to the first queue based on the registered data requirements.

18. The method of claim 17, wherein the step of registering further comprises registering priority requests by at least one of the plurality of consumers.

19. The method of claim 17, wherein the step of managing comprises the step of maintaining an IN pointer for data storage in the first queue and one OUT pointer for each

of the registered plurality of consumers to determine the amount of the data flow in and out of the first queue.

20. The method of claim 17, further comprising the step of prioritizing data consumption of the first queue based on an amount of unread data in the first queue of each of the plurality of consumers.

21. The method of claim 16, wherein the method is implemented in an embedded engine.

22. The method of claim 16, wherein the method is implemented in a telephony system.

23. A program storage device readable by a machine, tangibly embodying a program of instructions executable by the machine to perform method steps for sharing data between multiple consumers, the method comprising the steps of:

storing data in a first queue;

sharing the data in the first queue between a plurality of consumers; and

managing the storage of the data in the first queue and the consumption of the data in the first queue by each of the plurality of consumers.



24. The program storage device of claim 23, wherein the instructions for performing the step of sharing comprise instructions for performing the steps of:

5 registering data requirements of each of the plurality of consumers; and

assigning the plurality of consumers to the first queue based on the registered data requirements.

25. The program storage device of 24, wherein the instructions for performing the step of registering further  
10 comprise instructions for performing the step of registering priority requests by at least one of the plurality of consumers.

26. The program storage device of 24, wherein the instructions for performing the step of managing comprise  
15 instructions for performing the step of maintaining an IN pointer for data storage in the first queue and one OUT pointer for each of the registered plurality of consumers to determine the amount of the data flow in and out of the first queue.

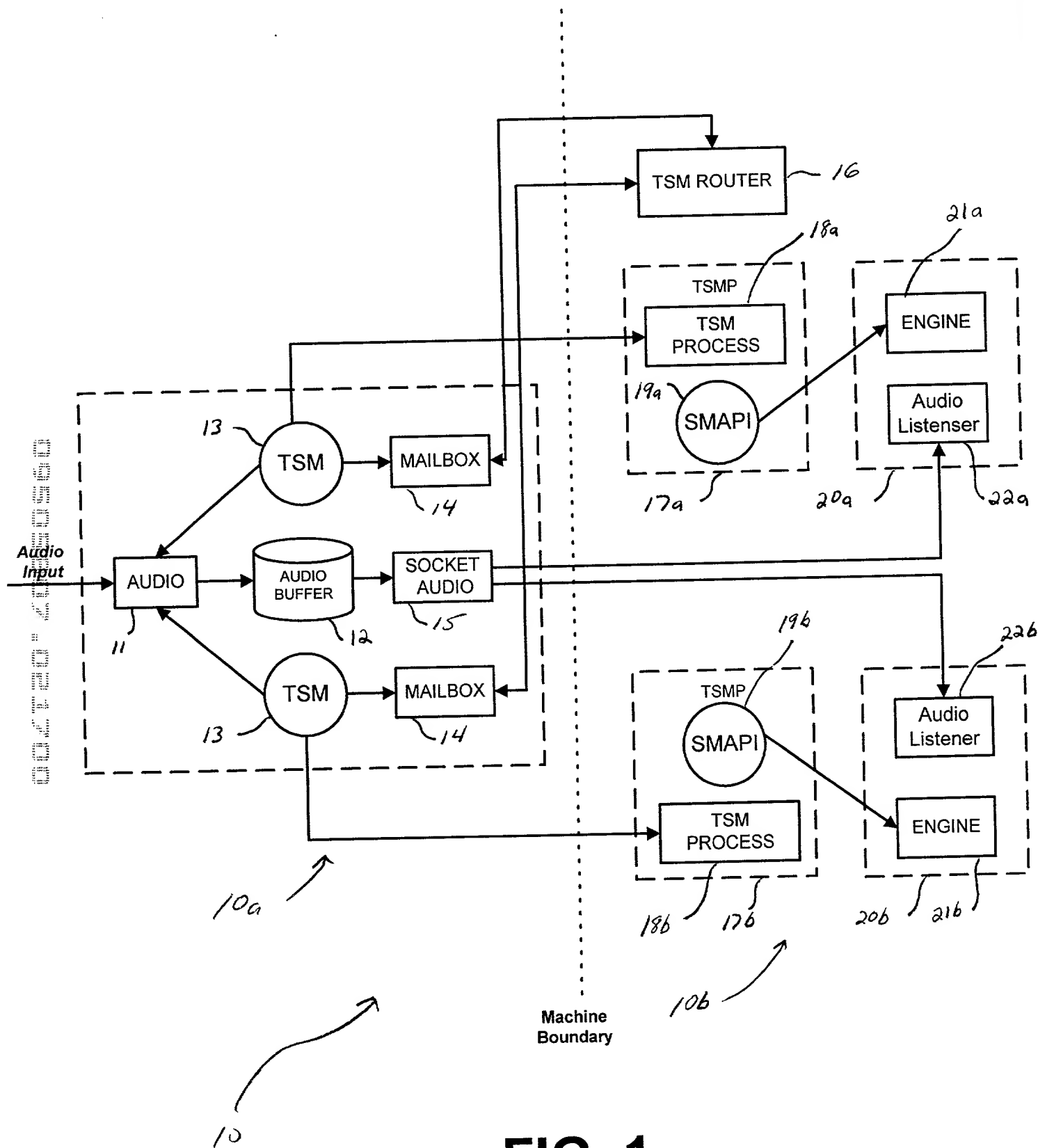
20

27. The program storage device of claim 24, further comprising instructions for performing the step of prioritizing data consumption of the first queue based on an amount of unread data of each of the plurality of consumers.

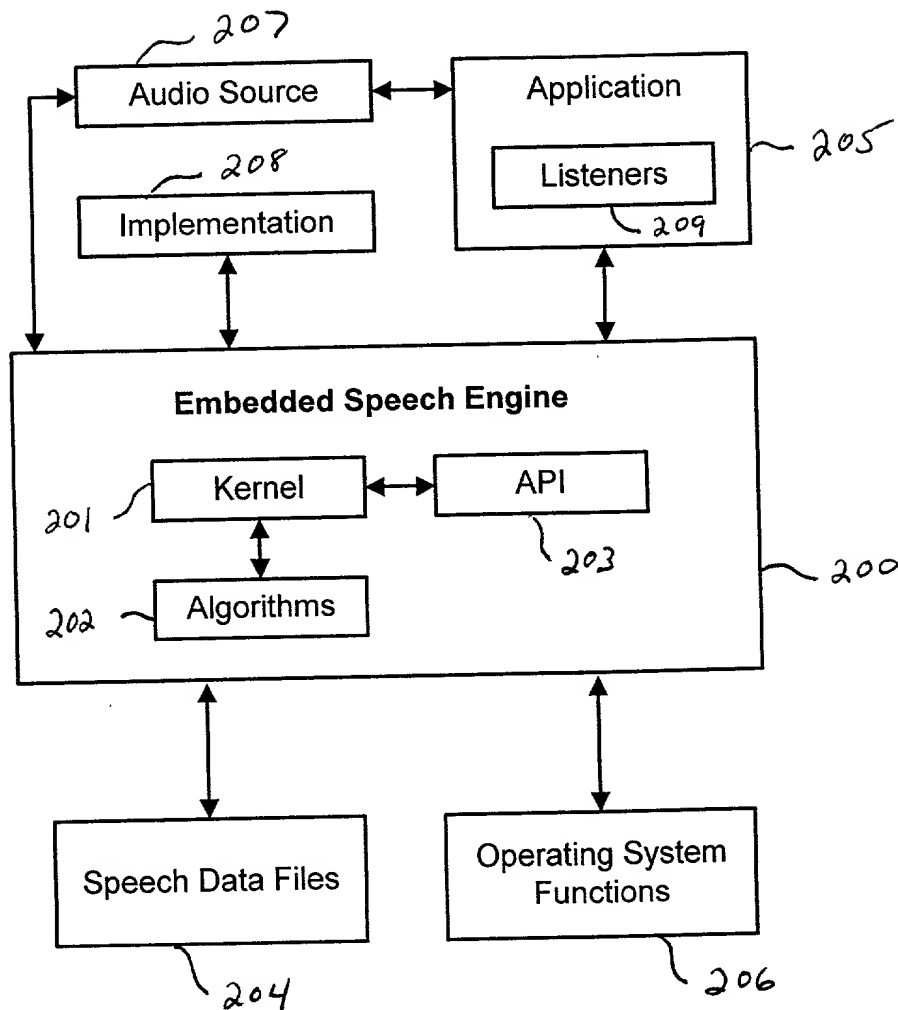
**SYSTEMS AND METHODS FOR PROCESSING AUDIO  
USING MULTIPLE SPEECH TECHNOLOGIES**

**ABSTRACT OF THE DISCLOSURE**

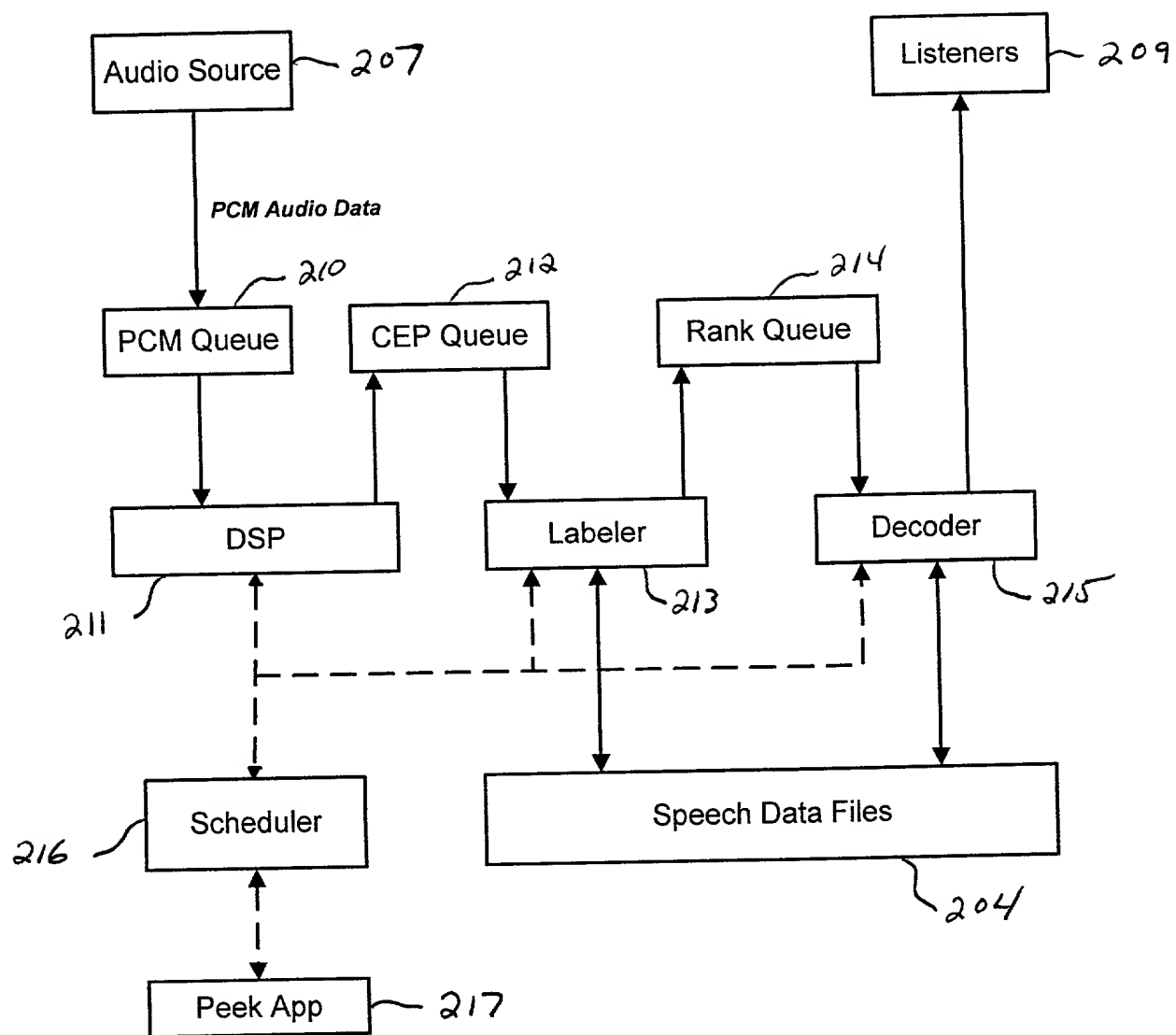
An audio splitting system for sharing speech data associated with the same utterance between multiple speech technologies (consumers) comprises a first queue for storing data; a plurality of consumers each sharing the data stored in the first queue; and a scheduler for managing the storage of the data in the first queue and the consumption of the data in the first queue by each of the plurality of consumers. In another aspect, the system comprises a plurality of queues and plurality of consumers. The consumers may include speech engines such as feature extraction engines, speech decoding engines, and speaker identification/verification engines, as well as data compression and decompression engines. The consumers will register their data requirements and priority requests with the scheduler. The scheduler assigns each of the plurality of consumers to one or more of the plurality of queues based on the registered data requirements. In this manner, the sharing of audio data (i.e., audio splitting) can occur at different stages in an I/O processing chain by, e.g., distributing digitized waveforms between different consuming engines and distributing features obtained at several stages of processing of the audio stream.



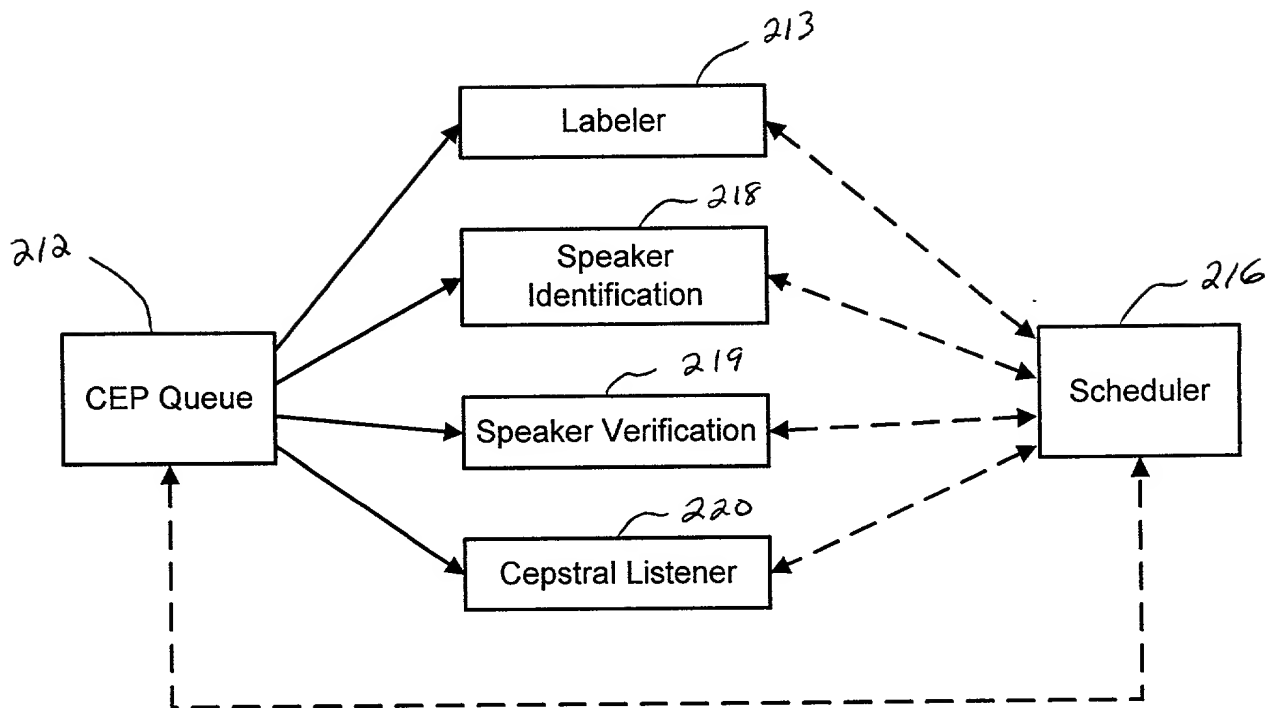
**FIG. 1**



**FIG. 2**



**FIG. 3**



**FIG. 4**

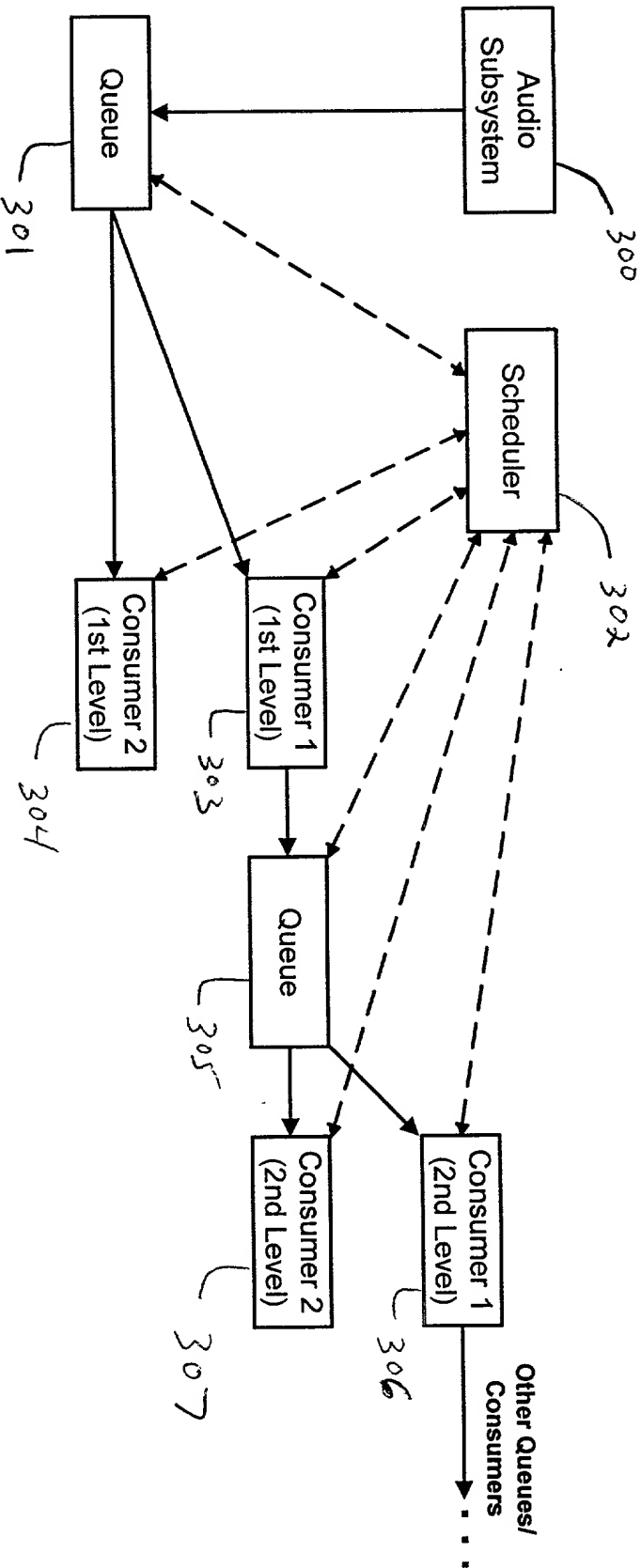


FIG. 5



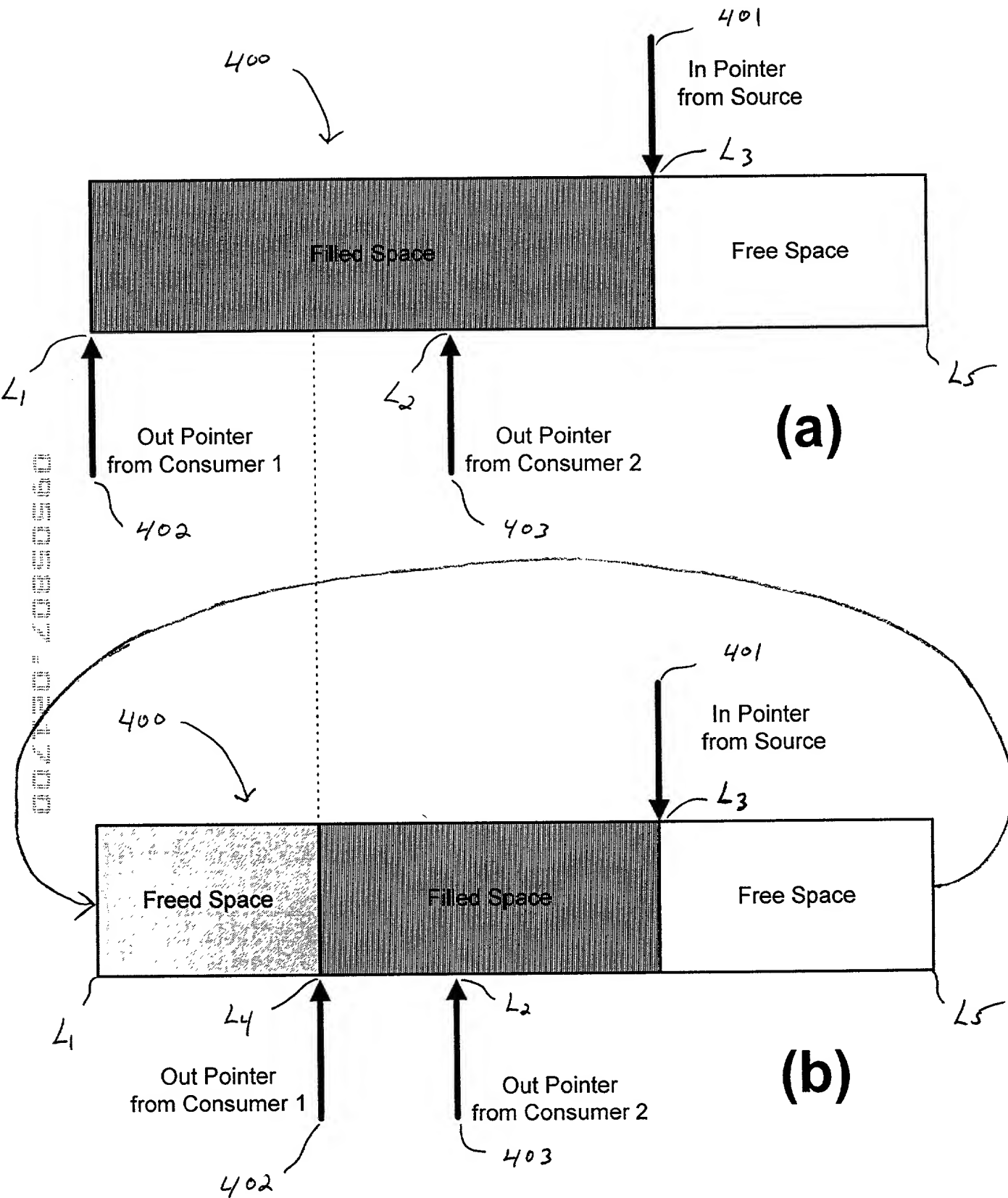


FIG. 6

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

**APPLICANT(S):** Abraham P. Ittycheriah, Stephane H. Maes,  
Jan Sedivy

**SERIAL NO.:** Unassigned

**FILED:** Concurrently herewith

**FOR:** **SYSTEMS AND METHODS FOR PROCESSING AUDIO  
USING MULTIPLE SPEECH TECHNOLOGIES**


**ASSOCIATE POWER OF ATTORNEY**

Please recognize **FRANK CHAU**, Reg. No. 34,136; **JAMES J. BITETTO**, Reg. No. 40,513; **FRANK V. DeROSA**, Reg. No. 43,584; and **GASPARE J. RANDAZZO**, Reg. No. 41,528; each of them of **F. CHAU & ASSOCIATES, LLP**, 1900 Hempstead Turnpike, Suite 501, East Meadow, New York 11554 as associate attorneys in the above-mentioned application, with full power to prosecute said application, to make alterations and amendments therein, and to transact all business in the Patent and Trademark Office connected therewith.

Telephone calls should be made to Frank Chau by dialing (516) 357-0091.

All written communications are to be sent to Frank Chau, Esq.,  
**F. Chau & Associates, LLP**, 1900 Hempstead Turnpike, Suite 501, East Meadow,  
New York 11554.

International Business Machines  
Corporation  
T.J. Watson Research Center  
Route 134 and Kitchawan Road  
Yorktown Heights, New York 10598

  
Manny W. Schecter  
Registration No. 31,722  
Paul J. Otterstedt  
Registration No. 37,411  
Attorney for Applicant(s)